AD-A235 514

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

3. REPORT TYPE AND DATES COVERED: FINAL 01 Sep 87 TO 30 Sep 90

4. TITLE AND SUBTITLE

Connectionist Models for Intelligent Computation

5. FUNDING NUMBERS

2305/B1

6. AUTHOR(S)

Professor Chen

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Maryland
Department of Physics
College Park Md 20742

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR/NE
Bldg 410
Bolling AFB Washington DC 20332-6448
Dr. Alan E. Craig

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AFOSR-87-0388

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

SEE REPORT FOR ABSTRACT

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

DTIC COPY INSPECTED 4

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. Z39-18

91 4 26 025

TECHNICAL REPORT

CONNECTIONIST MODELS FOR INTELLIGENT COMPUTATION, AFOSR-87-0388

Dr. H.H. Chen, Principal Investigator
Dr. Y.C. Lee, Co-Principal Investigator

II. REPORT OF PROGRESS

In the past few years, we have worked on various problems concerning neural networks. We list them below with brief descriptions of their significance.

a. Neural Networks With High Order Connections

A neural network derive its power of memory and learning from its connections. Perception is the simplest neural network that has a single layer of binary weights that connect each input neuron to an output neuron. The processing power of a perceptron is limited primary because of the binary nature of its connections. The spatial correlations of the input patterns are not utilized for the task that the perceptron is asked to do. There are several ways to add the spatial correlations into a network. One popular approach is to use hidden units in a multilayered net. The use of back-propagation algorithm to train such feed-forward networks has become wide spread. One of the disadvantage of this approach is the excruciatingly slow learning that the backprop incurs. This situation can be improved dramatically in many cases that the involved nonlinearity can be sort of inferred from the analysis of the problem. We could then use the single layer topology with higher order connections. A re-evaluation of the higher order network carried out by our group suggests many powerful techniques to reduce the number of spurious connections,which was the main obstacle to wide application of higher order networks. Prior-knowledge of the problem and the symmetry invariances etc. can be utilized efficiently to simplify the architecture complexity, to increase significantly the capacity of the associate memory, to reduce significantly the training time required to carry out the specific task in mind.

Many such examples had been studied. One is a landmark learning problem that has been studied by Barto, Anderson and Sutton. A bug is trained to direct its movement toward a tree aided by the landmarks in the environment. The higher order network learned the problem faster than the binary network by two orders of magnitude. Another problem deals with the storage capacity of higher order Hopfield net. We established that the higher order net could store orders of magnitude higher numbers of patterns in networks with, say, second order connections than that with first order connection. Other examples include the stereopsis network that necessarily require higher order connections to correlate the images that the left and the right eyes see, and the grammatical inference problems that involve a neural network finite state controller that is most naturally represented by a higher order recurrent net.

b. **Learning Stereopsis with Neural Networks**

Neural network models are very effective in dealing with perceptive problems such as vision, speech and motor control. One of the most prominent advantage of the neural network approach is the ability for it to automatically acquire the program from learning. On the other hand, since the learning process is usually very tedious and numerically intensive, it is usually difficult to make sense out of the acquired weights and make theoretical analysis about them. In this work, we have succeeded in training a higher order network analytically to perform stereopsis on random dot stereogram. The analytically calculated weights, obtained through the Hebbian learning rule, rediscovered the uniqueness and the continuity constraint proposed by Marrs and Poggio.

c. **Efficient Learning Algorithm for Neural Networks.**

Most neural networks possess a huge number of parameters to be adjusted while they are also being presented with an inordinate amount of patterns during training. These characteristics pose serious problems for the conventional optimization algorithm. Highly optimized conventional scheme such as the Newton's method requires too much storage and computations for problems having more than 100 adjustable parameters. On the other hand, the memory efficient conjugate gradient scheme has difficulty to handle a continuous stream of input data. The recursive least mean square method is on-line and provides quadratic convergence requires however N^2 operations and is applicable only to 'linear' parameters. The stochastic gradient descent seems to be the natural choice to deal with these problems but is very slow and is hampered by the 'ravine' problem.

To attack these problems, we have studied the high order stochastic gradient descent algorithm. The Hinton's empirical 'momentum' term is an example of the second order stochastic gradient descent method.However, because of its empirical nature, it is far from optimal both in terms of speed and convergence. Our work indicates that the average convergence rate for an *n-th* order stochastic gradient descent method is proportional to $(\lambda_1 / \lambda_N)^{1/n}$, where λ_1 and λ_N are the smallest and the largest eigenvalue of the average Hessian matrix, respectively. Since the condition number λ_1 / λ_N is typically a very small number (the ravine phenomena), the higher order scheme clearly represents a drastic improvement in the speed of convergence.

d. **Parallel Sequential Induction Network**

Most of the neural network research paid attention to improve the efficiency of learning algorithms with a fixed topology. In contrast, little progress has been made toward uncovering the designing principles for an optimal network topology. One plausible solution for the above problem was called a 'Parallel Sequential Induction Network'. As the name suggests, it combined the best of both the parallel and the sequential strategies to optimize the performance of a neural network classifier.The network first take the parallel approach by assigning an output decision neuron to each local decision region in the pattern space. Instead of letting a single decision neuron to carry the full burden of figuring out the full complex decision all by itself, the many decision neurons would share that responsibility and the individual task (part of the complex decision boundary) would be much simpler. The role of these local decision neurons are in a sense very similar to that

of the hidden neurons in a multi-layered net. The crucial difference is that our decision neurons are not hidden. Their connection weights are therefore much easier to train. Furthermore, the category label of these neurons are determined by a self-organization principle and are not supervised directly. We use an entropy measure that reflects the purity of patterns that were channelled to a node If the first layer of decision neurons are insufficient in completely classifying the patterns, we can always add another layer of descendent neurons to fine tune the result. The above combination of the parallel and the sequential strategies would enable us to shape the topology of a network automatically for an optimal performance in classifying patterns.

e. Higher Order Recurrent Networks and Grammatical Inference

Biological networks readily and easily process temporal information; artificial neural networks should do the same. Recurrent neural network models permit the encoding and learning of temporal sequences. The successive states of the system are encoded as the activity patterns of the neurons in a recurrent network. sequential input would cause the system state to make transitions from one to the other. A formal model of sequences that machine can generate and recognize is the formal grammar hierarchy that Chomsky classified. The simplest level of complexity is defined by a finite state machine and its associated regular grammar. The next level of complexity is described by pushdown automata and their associated context free grammars. The pushdown automata is a finite state machine with the added power to use a stack memory. Simple grammatical inference is defined as the problem of finding (learning) a grammar from a finite set of symbol string samples. In the context of a neural network, the grammatical inference is defined as the task of learning the machine that recognizes the teaching and the testing samples.

There has been many attempts in teaching neural nets to recognize grammars and simulate automata. However, as far as we know, nobody has studied systematically the grammatical inference at all levels of complexity. For example, Allen had attempted to learn some context free grammars using only a recurrent neural net without a stack memory. The result is that the neural net can not learn the grammar for strings with a length exceeding, say, five symbols. We used a higher order connection in the recurrent network which we showed to be sufficient in representing any given automata, and also devised a novel soft stack memory so that the neural net controller can be taught to use it. The result is that the neural network pushdown automata is being trained successfully without any prior knowledge or heuristics to recognize perfectly a few very important examples of context free grammars such as the parenthesis checker and the palindrome. More complex grammars such as the context sensitive grammar would need the power of a Turing machine to recognize them. In our neural network approach, what is needed is a tape memory that the neural network can be trained to read, write, or erase information on it. This is a much more difficult task and will be tackled in the near future. However, since any one dimensional tape can be decomposed into two stacks, It seems plausible that we could transfer our knowledge of training a neural network to use the stack to the use of a memory tape.The ability of the neural network to extract complex grammatical rules from examples of sequential patterns is a very important step toward the understanding of higher level reasonings that still eludes us in the quest of understanding the human intelligence.